Inverse Problem with Mapper Graphs
report

Course: Projects in Topological Data Analysis HS24

Joint initiative: Iowa State University, Vassar College, University of Notre Dame, University of Oregon, TU Eindhoven, University of Fribourg, ETH Zürich

Supervisors: Dr Robin Belton (Vassar College) and Dr Enrique Alvarado (Iowa State University)

Author: Maja Gwóźdź (ETH Zürich)

Date: December 18, 2024

Acknowledgments
I would like to express sincere gratitude to my Supervisors. In particular, I want to thank them for creating a welcoming atmosphere, providing helpful guidance, and making the course very enjoyable. Last but not least, I also want to thank Dr Patrick Schnider for the seamless organisation and coordination.

1 Introduction

We present a heuristic[1] for the construction of convex sets, either in $\mathbb{R}^3$ or $\mathbb{R}^6$, whose nerve matches the input graph G. As a matter of fact, the convex sets will be balls of possibly different radii. We emphasise that this is a heuristic, not an algorithm with theoretical guarantees, due to the fact that our approach relies on randomised initialisation and iterative optimisation. The method proposed in this note is a direct extension of the results from [1], where the authors affirmatively answered the question whether for some dataset X and a graph G, there exists a set of Mapper parameters such that the output Mapper graph of X is isomorphic to G. In their proof, they crucially rely on a result of Wegner [20] – Perel'man [12] rediscovered this result in 1985 – which states that any d-dimensional simplicial complex is isomorphic to the nerve of a collection of convex sets in $\mathbb{R}^{2d+1}$. A sketch of the proof is provided in [15, Theorem 3.1]. Although the proof is constructive, the convex sets are not balls.

[1]The full code is available at: https://github.com/mkg33/PiTDA/blob/main/tda_final.py. Consult this page for more sample outputs: https://github.com/mkg33/PiTDA/tree/main/output.

The following method is capable of processing both sparse and large graphs, even if the number of nodes exceeds 150. The execution times are not prohibitive, even for such inputs. The value of our contribution lies more in the combination of existing computational techniques, rather than in providing a completely novel approach to the construction of graph embeddings or, in particular, convex sets.

2 Related works

The topic of graph embeddings has a rich literature (*cf.* [18]). Our method has been predominantly inspired by several known facts, most notably, the simple observation that every k-dimensional simplicial complex has a geometric realization in $\mathbb{R}^{2k+1}$ ([14], *cf.* [4]). Indeed, it suffices to place the vertices as distinct points on a moment curve. This is precisely why our heuristic attempts to construct the embedding in $\mathbb{R}^6$ in the initial step. We have also been inspired by certain graph drawing methods. For instance, the algorithm proposed by Kamada and Kawai on drawing undirected graphs [7], especially drawings conforming to certain geometric constraints, such as lying on concentric circles.

Another relevant domain in this context is intersection graph theory. An intersection graph is understood as a graph representing the intersection pattern of a set family. In our work, the basic insights from [10] had an impact on our handling of overlaps (see below). In a similar vein, the results from [19] on contact (tangency) graphs helped us in the construction of the dissimilarity matrix.

The circle packing theorem (also known as the Koebe-Andreev-Thurston theorem [17]) is conceptually closely related to our approach. The theorem states that: For a (finite) maximal planar graph G, the circle packing whose tangency graph is isomorphic to G is unique, up to transformations. We aim to use the constructive proof in our subsequent research.

3 Method

In the preprocessing stage, we simply convert the input graph to an adjacency list, which facilitates the computation. The rest of the pipeline could be summarised as follows. Our method takes the input graph (as an adjacency list) and embeds its vertices into $\mathbb{R}^6$ via multidimensional scaling (MDS)[2] in the initial step. More precisely, we construct a dissimilarity matrix with entries encoded as follows: if a given pair of edges is connected, we put 1 in the respective entry, and 0 otherwise. We have also experimented with PCA but it yielded significantly worse results. We note in passing that classical

[2]Also known as *Principal Coordinates Analysis*. For a comprehensive treatment, consult [3].

MDS is equivalent to PCA provided that Euclidean distance is used [6]. Intuitively, MDS works by preserving pairwise distances or dissimilarities between points in a reduced space.

In the next step, we apply a nonlinear optimisation method, namely, *Sequential Least Squares Programming* (SQLSP) (*cf.* [8] and [11]), which we have selected on the basis of preliminary experimental results. It is conceivable that other optimisers would match our baseline performance, and we intend to explore this question in future research. The nonlinear optimiser is used to position the balls so that connected components correspond to overlaps or contacts. If there is no shared edge, the balls are disjoint, as expected. The objective function minimises the sum of assigned radii and centres of the balls.

By default, the optimiser runs iteratively for a maximum of five attempts. If the optimiser succeeds at finding balls that respect the aforementioned constraints, we obtain convex sets in $\mathbb{R}^6$. In the case of planar input graphs, the above method is executed again in order to lower the final dimensionality to 3. If the optimiser does not fail at this stage, the output consists of convex sets in $\mathbb{R}^3$. Finally, the method then constructs a nerve complex from the resulting convex sets (by means of `Gudhi`'s functionality [9]), and verifies correctness by comparing it against the original graph.

For completeness, we state the nerve theorem (*cf.* [2], [5], [16], [14]) below:

Theorem 1 (Nerve Theorem) *Let $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ be a collection of open subsets of a topological space X such that:*

1. *Each non-empty finite intersection $U_{\alpha_0} \cap U_{\alpha_1} \cap \cdots \cap U_{\alpha_k}$ is either empty or contractible.*
2. *The union of the collection satisfies $\bigcup_{\alpha \in A} U_\alpha = X$.*

Define the nerve of $\mathcal{U}$, denoted $\mathcal{N}(\mathcal{U})$, as the simplicial complex with:

- *Vertices corresponding to the elements of $\mathcal{U}$.*
- *A k-simplex for every non-empty $k+1$-fold intersection $U_{\alpha_0} \cap U_{\alpha_1} \cap \cdots \cap U_{\alpha_k} \neq \emptyset$.*

Then, there exists a homotopy equivalence between X and the geometric realization $|\mathcal{N}(\mathcal{U})|$ of the nerve $\mathcal{N}(\mathcal{U})$.

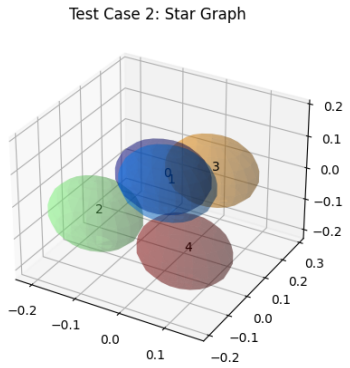
We also provide utility functions for the visualisation of the convex sets (only for planar graphs) and for reporting detailed results, such as execution times and cactus plots juxtaposing execution times against the number of nodes. To allow for more comprehensive and transparent testing, we include a function that generates random regular graphs with specified parameters, such as graph degree and the number of nodes. Additionally, the code contains a simple function for generating random trees,

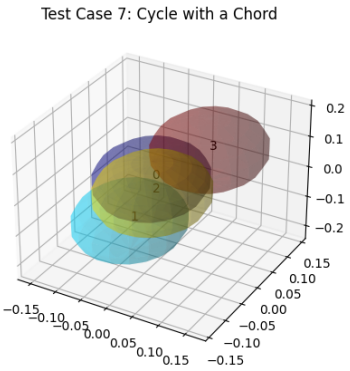
based on the Prüfer method [13]. For practical reasons, there exists a flag that lets the user abort the test once a timeout condition has been reached.

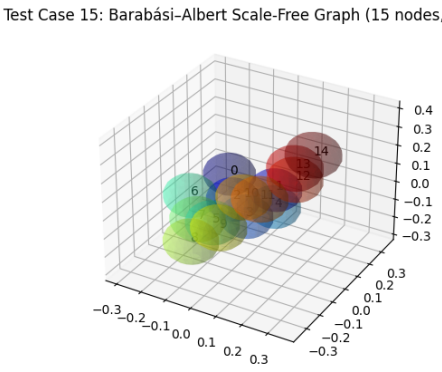
4 Results

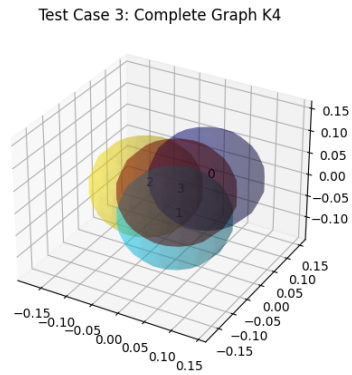
We have performed over 1200 tests on a heterogeneous set of graphs. The examples included, *inter alia*, standard textbook examples (K_4, graphs with chords, a Windmill graph, *etc.*) and many random regular graphs with varying parameters (graph degree and number of nodes). Indeed, the vast majority of test cases included random graphs – only 60 test cases have been manually hard-coded. In the case of random regular graphs, we used degrees of: 2, 4, 6, and 8. All test cases have passed, albeit with varying execution times. We set the timeout parameter to 3 hours, so it is conceivable that graphs with, say, over 250 nodes would eventually lead to failure. For the first 60 "textbook" test cases, the total execution time (rounded up to the nearest decimal place) was 534.6 seconds, with a mean of 8.9 seconds. For 101 cases of random regular graphs with an increasing number of nodes in each iteration, the total was 10558.9 seconds, with a mean of 104.5 seconds.

It is particularly interesting to note that the optimiser never exceeded the default number of attempts (5), and that it reached at most 4 in very rare cases, while still being able to find the required solution. Below are sample visualisations of the resulting convex sets:









Let us now examine selected results in more detail. All experiments have been run on a CPU – MacBook Air with an M1 chip.

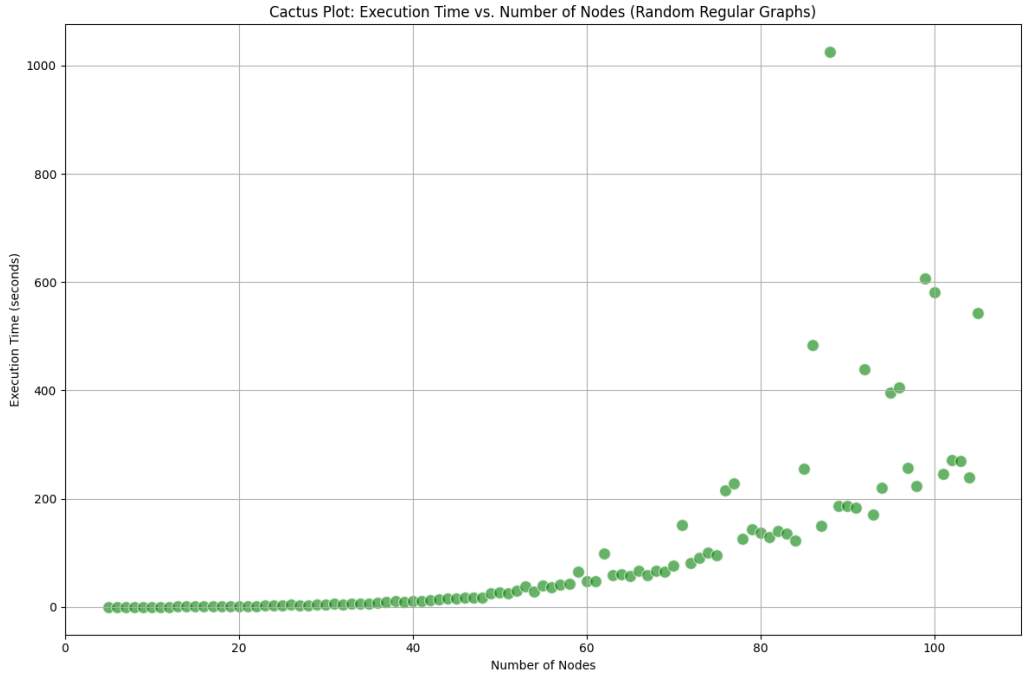


Figure 1: Cactus plot showing the relationship between the execution time (CPU) and the number of nodes in the input random regular graphs. Results from one trial run, with graph degree fixed at 4.

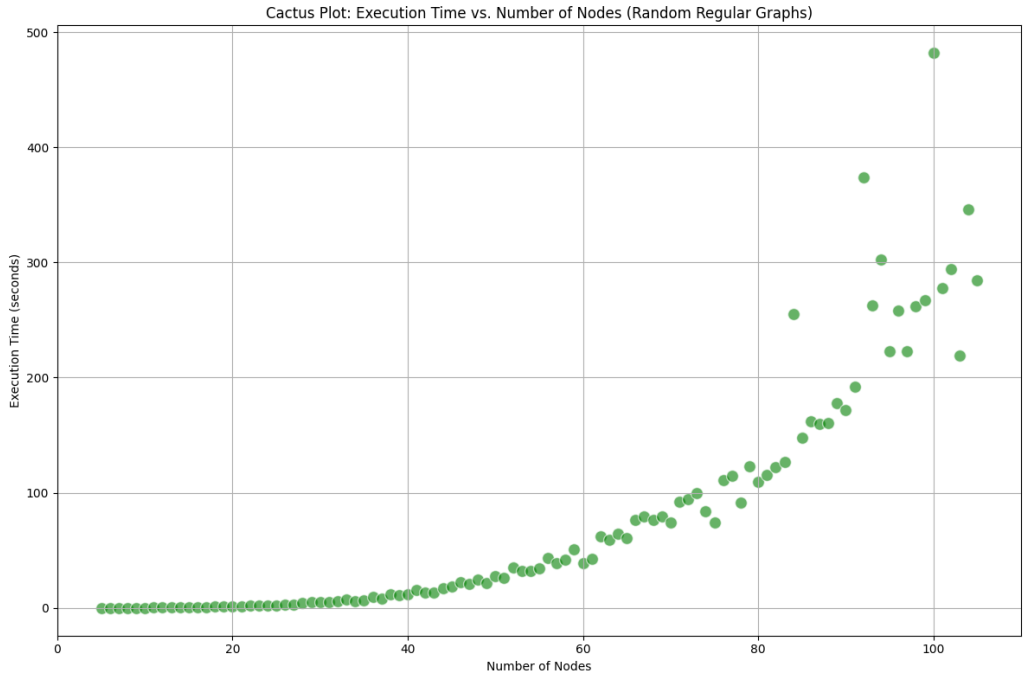


Figure 2: Cactus plot showing the relationship between the execution time (CPU) and the number of nodes in the input random regular graphs. Results from one trial run, with graph degree fixed at 2.

The above results clearly indicate that the method's performance is satisfactory. In the case of random regular graphs of degree 4, the execution time begins to increase significantly around 80 nodes, while for degree 2 it occurs already at *ca* 60 nodes. This phenomenon can be explained by observing that the lesser degree leads to more intersection restrictions.

5 Conclusions

We have presented a heuristic for constructing convex sets (in our context, balls) in $\mathbb{R}^3$ or $\mathbb{R}^6$ (depending on the planarity of the input graph) whose nerve matches the input graph. The procedure combined several known methods, such as MDS or SQLSP, and performed surprisingly well on an extensive test set.

The present project served to establish a working method and to collect enough experimental evidence. In subsequent analyses, we aim to provide concrete bounds on the heuristic and to formalise our idea. Future research directions will encompass reduced reliance on the planarity condition, *i.e.,* we are aware of the existence of non-planar graphs that can be embedded in 3 dimensions, so our current embedding of non-planar graphs in $\mathbb{R}^6$ is an oversimplification. The fact that planar graphs are initially positioned in $\mathbb{R}^6$ and then iteratively projected to $\mathbb{R}^3$ further contributes to the overall computational overhead. Preliminary tests with $\mathbb{R}^5$ and even $\mathbb{R}^4$ demonstrate that significant improvements are possible.

As we have already noted above, other nonlinear optimisers might match the baseline performance, while being more computationally efficient. A comprehensive comparison is required to test this hypothesis. Furthermore, it is important to construct more test cases, especially those that might be edge cases. The current testing procedure relies predominantly on random graphs or trees restricted to a small range of degrees.

The above limitations notwithstanding, we believe that the proposed method still serves as a useful heuristic, which is corroborated by the experimental results. We hope that our approach will prove useful in exploratory research on graph theory or combinatorial topology.

[1] Enrique G Alvarado, Robin Belton, Kang-Ju Lee, Sourabh Palande, Sarah Percival, Emilie Purvine, and Sarah Tymochko. Any graph is a mapper graph. *arXiv preprint arXiv:2408.11180*, 2024.

[2] Ulrich Bauer, Michael Kerber, Fabian Roll, and Alexander Rolle. A unified view on the functorial nerve theorem and its variations. *Expositiones Mathematicae*, 41(4):125503, 2023.

[3] Ingwer Borg and Patrick J.F. Groenen. *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media, 2007.

[4] Herbert Edelsbrunner. *Algorithms in combinatorial geometry*, volume 10. Springer Science & Business Media, 1987.

[5] Herbert Edelsbrunner and John L Harer. *Computational topology: an introduction*. American Mathematical Society, 2022.

[6] Benyamin Ghojogh, Mark Crowley, Fakhri Karray, and Ali Ghodsi. Multidimensional scaling, sammon mapping, and isomap. In *Elements of Dimensionality Reduction and Manifold Learning*, pages 185–205. Springer, 2023.

[7] Tomihisa Kamada, Satoru Kawai, et al. An algorithm for drawing general undirected graphs. *Information processing letters*, 31(1):7–15, 1989.

[8] Dieter Kraft. A software package for sequential quadratic programming. *Forschungsbericht-Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt*, 1988.

[9] Clément Maria, Jean-Daniel Boissonnat, Marc Glisse, and Mariette Yvinec. The Gudhi library: Simplicial complexes and persistent homology. In *Mathematical Software–ICMS 2014: 4th International Congress, Seoul, South Korea, August 5-9, 2014. Proceedings 4*, pages 167–174. Springer, 2014.

[10] Terry A McKee and Fred R McMorris. *Topics in intersection graph theory*. SIAM, 1999.

[11] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 1999.

[12] Grigori Ya Perel'man. Realization of abstract k-skeletons of intersections of convex polyhedra in $\mathbb{R}^{2k-1}$. *Geometric Questions in the Theory of Functions and Sets*, pages 129–131, 1985.

[13] Heinz Prüfer. Neuer Beweis eines Satzes über Permutationen. *Archiv derMathematik und Physik*, 27:742–744, 1918.

[14] Patrick Schnider and Simon Weber. Introduction to topological data analysis. ETH Zürich, 2024. Lecture Notes.

[15] Martin Tancer. *Intersection Patterns of Convex Sets via Simplicial Complexes: A Survey.*, pages 521–540. Springer New York, 2013.

[16] Martin Tancer and Dmitry Tonkonog. Nerves of good covers are algorithmically unrecognizable. *SIAM Journal on Computing*, 42(4):1697–1719, 2013.

[17] William P Thurston. *The geometry and topology of three-manifolds: With a preface by Steven P. Kerckhoff*, volume 27. American Mathematical Society, 2022.

[18] Csaba D Toth, Joseph O'Rourke, and Jacob E Goodman. *Handbook of discrete and computational geometry*. CRC press, 2017.

[19] Richard Wagner. Ein Kontaktproblem der konformen Abbildung. 1956.

[20] Gerd Wegner. Eigenschaften der Nerven Homologisch-einfacher Familien in $\mathbb{R}^n$. *Ph.D. thesis, Georg-August-Universität Göttingen*, 1967.